



Munich Personal RePEc Archive

# **Confidence intervals for percentiles in stationary ARMA processes: An application using environmental data**

George Halkos and Ilias Kevork

Department of Economics, University of Thessaly

May 2014

Online at <http://mpra.ub.uni-muenchen.de/56134/>

MPRA Paper No. 56134, posted 22. May 2014 13:19 UTC

# Confidence intervals for percentiles in stationary ARMA processes: An application using environmental data

George E. Halkos & Ilias S. Kevork

Laboratory of Operations Research,

Department of Economics, School of Humanities, University of Thessaly,

43 Korai Street, Volos 38333, Greece.

[halkos@uth.gr](mailto:halkos@uth.gr)

[kevork@uth.gr](mailto:kevork@uth.gr)

## Abstract

Percentiles estimation plays an important role at the stage of making decisions in many scientific fields. However, the up-to-now research on developing estimation methods for percentiles has been based on the assumption that the data in the sample are formed independently. In the current paper we suppress this restrictive assumption by assuming that the values of the variable under study are formed according to the general linear process. After deriving the asymptotic distribution of the Maximum Likelihood estimator for the  $100 \times P_{th}$  percentile, we give the general form of the corresponding asymptotic confidence interval. Then, the performance of the estimated asymptotic confidence interval is evaluated in finite samples from the stationary AR(1) and ARMA(1,1) through Monte-Carlo simulations by computing two statistical criteria: (a) the actual confidence level, (b) the expected half-length as percentage of the true value of the percentile. Simulation results show that the validity of the estimated asymptotic confidence interval depends upon the sample size, the size of the 1<sup>st</sup> order theoretical autocorrelation coefficient, and the true cumulative probability  $P$  related to the percentile. Finally, an application example is given using the series of the CO<sub>2</sub> annual emissions intensity in Greece (kg per kg of oil equivalent energy use) for the period 1961-2010. Confidence intervals for percentiles are constructed on this series and discussion about the validity of the estimation procedure follows according to the findings from the simulation experiments regarding the values of the aforementioned criteria.

**Keywords:** Percentiles; environmental data; time series models; confidence intervals.

**JEL Classification Codes:** C13; C22; C53; Q50; Q54.

Work in this study has received funding from the "GHGsMETI" program, which takes place within the SYNERGASIA 2011 action and is supported by the European Regional Development fund and Greek National Funds, project number 11SYN\_8\_118. The text represents the authors' views.



**Ministry of Education and Religious Affairs**  
**GSRT - Management and Implementation Agency for RTD and Innovation Activities**

O.P. Competitiveness and Entrepreneurship (EPAN II), ROP Macedonia - Thrace, ROP Crete and Aegean Islands, ROP Thessaly - Mainland Greece - Epirus, ROP Attica

# Διαστήματα εμπιστοσύνης για εκατοστημόρια σε στάσιμες ARMA διαδικασίες: Μία εμπειρική εφαρμογή σε περιβαλλοντικά δεδομένα

Ηλίας Κεβόρκ και Γεώργιος Χάλκος

Εργαστήριο Επιχειρησιακών Ερευνών

Τμήμα Οικονομικών Επιστημών, Σχολής Κοινωνικών και Ανθρωπιστικών Επιστημών,  
Πανεπιστήμιο Θεσσαλίας, Κοραή 43, Βόλος 38333

[halkos@uth.gr](mailto:halkos@uth.gr)

[kevork@uth.gr](mailto:kevork@uth.gr)

## Περίληψη

Η εκτίμηση εκατοστημορίων παίζει πλέον σημαντικό ρόλο στα διάφορα στάδια λήψης αποφάσεων σε πολλούς επιστημονικούς τομείς. Όμως η μέχρι τώρα έρευνα ανάπτυξης μεθόδων εκτίμησης των εκατοστημορίων βασίστηκε στην υπόθεση ότι οι παρατηρήσεις στο δείγμα διαμορφώνονται ανεξάρτητα μεταξύ τους. Στην παρούσα εργασία καταργούμε την υπόθεση αυτή υποθέτοντας ότι οι τιμές της υπό μελέτη μεταβλητής σχηματίζονται με βάση την γενική γραμμική στοχαστική ανάλυση. Εξάγοντας πρώτα την ασυμπτωτική κατανομή του εκτιμητή μεγίστης πιθανοφάνειας για το  $100 \times P_{th}$  εκατοστημόριο, δίνουμε στη συνέχεια τη γενική μορφή του αντίστοιχου ασυμπτωτικού διαστήματος εμπιστοσύνης. Η εγκυρότητα του διαστήματος αυτού όταν εκτιμάται σε στάσιμες σειρές AR(1) και ARMA(1,1) εξετάζεται μέσω χρήσης προσομοιώσεων Monte-Carlo και υπολογισμού δυο στατιστικών κριτηρίων: (α) του πραγματικού επιπέδου εμπιστοσύνης, και (β) του αναμενόμενου ημι-πλάτους του διαστήματος ως ποσοστό της πραγματικής τιμής του εκατοστημορίου. Τα αποτελέσματα των πειραμάτων προσομοίωσης δείχνουν ότι η εγκυρότητα του εκτιμηθέντος ασυμπτωτικού διαστήματος εμπιστοσύνης εξαρτάται από το μέγεθος του δείγματος, το μέγεθος του συντελεστή αυτοσυσχέτισης 1<sup>ης</sup> τάξης και την τιμή της αθροιστικής πιθανότητας P του εκατοστημορίου. Τέλος διενεργείται εφαρμογή της μεθοδολογίας εκτίμησης του ασυμπτωτικού διαστήματος εμπιστοσύνης στη σειρά *πυκνότητα των εκπομπών διοξειδίου του άνθρακα* στην Ελλάδα για τα έτη 1961-2010. Στην εφαρμογή αυτή σχολιάζεται η εγκυρότητα των εκτιμηθέντων διαστημάτων εμπιστοσύνης βάσει των ευρημάτων των πειραμάτων προσομοίωσης αναφορικά με τις τιμές των δύο παραπάνω στατιστικών κριτηρίων.

**Λέξεις Κλειδιά:** Εκατοστημόρια, περιβαλλοντικά δεδομένα, υποδείγματα χρονικών σειρών, διαστήματα εμπιστοσύνης.

**Κωδικοί JEL:** C13; C22; C53; Q50; Q54.

Η μελέτη αυτή έχει λάβει χρηματοδότηση από το πρόγραμμα «GHGsMETI», στα πλαίσια του έργου δράσης «ΣΥΝΕΡΓΑΣΙΑ 2011» με κωδικό αριθμό έργου 11SYN\_8\_118 και υποστηρίζεται από το Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης και από Ελληνικούς Εθνικούς Πόρους. Το κείμενο εκφράζει τις απόψεις των συγγραφέων.



ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ  
ΕΥΡΩΠΑΪΚΟ ΤΑΜΕΙΟ  
ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ



Υπουργείο Παιδείας και Θρησκευμάτων  
ΕΥΔΕ-ΕΤΑΚ



Ε. Π. Ανταγωνιστικότητα και Επιχειρηματικότητα (ΕΠΑΝ II), ΠΕΠ Μακεδονίας – Θράκης, ΠΕΠ Κρήτης και Νήσων Αιγαίου, ΠΕΠ Θεσσαλίας – Στερεάς Ελλάδας – Ηπείρου, ΠΕΠ Αττικής

## 1. ΕΙΣΑΓΩΓΗ

Η κατασκευή διαστημάτων εμπιστοσύνης για εκατοστημόρια αποτελεί πλέον θεματική έρευνας με αυξανόμενο ενδιαφέρον. Ήδη έχει αναγνωρισθεί στη διεθνή βιβλιογραφία η αναγκαιότητα της εκτίμησης εκατοστημορίων σε σειρές μεγεθών που αφορούν διάφορες κοινωνικοοικονομικές μεταβλητές όπως το οικογενειακό εισόδημα, η εξέλιξη του βάρους και τους ύψους των νεογέννητων παιδιών κλπ. Επιπλέον, οι εκτιμήσεις των εκατοστημορίων παίζουν σημαντικό ρόλο σε αποφάσεις που πρέπει να ληφθούν στην άσκηση επιχειρηματικής πολιτικής. Ως παραδείγματα αναφέρουμε τον προσδιορισμό των ασφαλιστρών στο κλάδο ασφαλειών, τη μέτρηση της εγκυρότητας σε θέματα μηχανικής, τον προσδιορισμό της ποσότητας παραγγελίας ή του σημείου αναπαραγγελίας σε υποδείγματα αποθεματικής πολιτικής newsvendor και υποδείγματα συνεχούς επιθεώρησης, και τον προσδιορισμό της συσσωρευτικής ικανότητας (assimilative capacity) σε υποδείγματα διαχείρισης του περιβάλλοντος. Ειδικότερα, στον τομέα του περιβάλλοντος, είναι πολύ σημαντικό να αναλυθούν οι επιπτώσεις των διαφόρων περιβαλλοντικών πολιτικών σε ξεχωριστά εκατοστημόρια της οριακής κατανομής της υπό εξέταση περιβαλλοντικής μεταβλητής αποφεύγοντας τα προβλήματα που ανακύπτουν από τη χρήση του μέσου αριθμητικού ως μοναδικής στατιστικής παραμέτρου.

Ένας σημαντικός αριθμός εργασιών έχει ήδη εμφανιστεί στη διεθνή βιβλιογραφία παρουσιάζοντας διαδικασίες στατιστικών ελέγχων και μεθόδους κατασκευής διαστημάτων εμπιστοσύνης για εκατοστημόρια όταν η πληθυσμιακή κατανομή για την υπό μελέτη μεταβλητή είναι άγνωστη. Η πιο σύνηθης προσέγγιση είναι η θεμελίωση απαραμετρικών (distribution-free) διαστημάτων εμπιστοσύνης χρησιμοποιώντας τη σχέση μεταξύ εκατοστημορίων και διάταξης (ordering) και κατάταξης (ranking) των παρατηρήσεων (Gibbons and Chakraborti, 2003. Chakraborti and Li, 2007). Από την άλλη πλευρά, η μέθοδος Bootstrapping αποτελεί σήμερα μια σημαντική εναλλακτική προσέγγιση ανάπτυξης

απαραμετρικών διαστημάτων εμπιστοσύνης για εκατοστημόρια εκμεταλευόμενοι τη δύναμη των ηλεκτρονικών υπολογιστών και τις δυνατότητες που δίνουν τα διάφορα εμπορικά υπολογιστικά πακέτα H/Y (Efron and Tibshirani, 1993). Επίσης, μια άλλη προσέγγιση εξαγωγής εκτιμητών για εκατοστημόρια βασίζεται στην αριστοποίηση μιας συνάρτησης απωλειών απόλυτων σφαλμάτων (absolute error loss function) χρησιμοποιώντας απαραμετρικές συναρτήσεις πυκνότητας πιθανότητας με το εκατοστημόριο να εκφράζεται ως το άθροισμα μιας παραμέτρου θέσης και το γινόμενο μιας σταθεράς επί μια παράμετρο κλίμακας (Keating 1983; Keating *et al.*, 2010).

Αντίθετα με τα παραπάνω, η εκτίμηση διαστημάτων επιστοσύνης για εκατοστημόρια υπό το πλαίσιο μιας παραμετρικής προσέγγισης έχει λάβει μέχρι τώρα μικρό ερευνητικό ενδιαφέρον. Πιο συγκεκριμένα, ένας μικρός αριθμός εργασιών έχει παρουσιαστεί στη διεθνή βιβλιογραφία υποθέτοντας ότι η κατανομή της υπό εξέταση μεταβλητής είναι η κανονική με άγνωστο μέσο και άγνωστη διακύμανση. Χρησιμοποιώντας το κριτήριο της αποτελεσματικότητας της εγγύτητας κατά Pitman (Pitman-closeness efficiency) οι Dyer *et al.* (1977) έκαναν σύγκριση διαφόρων εκτιμητών για εκατοστημόρια [μεταξύ των εκτιμητών που εξετασθηκαν ήταν οι Maximum Likelihood (ML), Minimum Variance Unbiased Estimator (MVUE), και Best Invariant Estimator (BIE)]. Οι Bland and Altman (1999) εξήγαγαν ένα συμμετρικό διάστημα εμπιστοσύνης για εκατοστημόρια χρησιμοποιώντας ιδιότητες των κατανομών δειγματοληψίας του δειγματικού μέσου και της δειγματικής διακύμανσης.

Στην έκταση της γνώσης μας επί του θέματος, οι Chakraborti and Li (2007) ήταν οι πρώτοι που διενήργησαν συγκρίσεις της εγκυρότητας μεταξύ διαφόρων μεθόδων εκτίμησης διαστημάτων εμπιστοσύνης για εκατοστημόρια, όταν η κατανομή της υπό εξέταση μεταβλητής είναι η κανονική με άγνωστο μέσο και άγνωστη διακύμανση. Εκτός των διαστημάτων εμπιστοσύνης τα οποία συνδέονται με τους εκτιμητές ML και MVUE, στην μελέτη αυτή συμπεριλήφθηκαν το διάστημα εμπιστοσύνης βασιζόμενο στον προσημικό

έλεγχο και το διάστημα πρόβλεψης της εκ των υστέρων Μπευνζιανής (Bayesian) ανάλυσης. Ως κριτήρια για τις συγκρίσεις αυτές, οι συγγραφείς χρησιμοποίησαν τις τιμές του πραγματικού επιπέδου εμπιστοσύνης (καλούμενο και ως κάλυψη) και του αναμενόμενου πλάτους του διαστήματος τις οποίες η κάθε μέθοδος επιτύγχανε. Βάσει των κριτηρίων αυτών οι συγγραφείς θεωρούν ως καλή επιλογή τη χρήση των διαστημάτων εμπιστοσύνης πεπερασμένων δειγμάτων που συνδέονται με τους εκτιμητές ML και MVUE, όταν για μεν το πρώτο εκτιμητή χρησιμοποιούνται οι κριτικές τιμές της student-t, ενώ για τον εκτιμητή ML οι κριτικές τιμές της non-central student-t.

Με τη κατανομή της υπό εξέταση μεταβλητής να είναι η κανονική με άγνωστο μέσο και άγνωστη διακύμανση, οι Donner και Zou (2010) παρουσίασαν μια μέθοδο κατασκευής ασυμμετρικών διαστημάτων εμπιστοσύνης χρησιμοποιώντας τη μέθοδο ανάκτησης εκτιμήσεων διακύμανσης (Variance Estimates Recovery). Επίσης στα πλαίσια εκτίμησης της άριστης ποσότητας παραγγελίας σε υποδείγματα αποθεματικής πολιτικής newsvendor ο Kevoork (2010) εξήγαγε το ασυμπτωτικό διάστημα εμπιστοσύνης του εκτιμητή ML. Τέλος, θεωρώντας ένα διαχρονικό δείγμα ζήτησης με την οριακή κατανομή να είναι η εκθετική, οι Halkos και Kevoork (2013a) εξήγαγαν εκτιμητή ο οποίος διασφαλίζει ότι η πιθανότητα η επόμενη χρονικά παρατήρηση να είναι μικρότερη της άριστης ποσότητας παραγγελίας να ισούται με την αθροιστική πιθανότητα P του εκατοστημορίου.

Η βασική όμως υπόθεση που έγινε σε όλες τις προαναφερθείσες εργασίες ήταν ότι οι παρατηρήσεις στο δείγμα διαμορφώνονται ανεξάρτητα μεταξύ τους. Παρόλα αυτά μπορεί να υπάρξουν περιπτώσεις μεταβλητών για τις οποίες η εξέλιξη των τιμών τους σε έναν συγκεκριμένο χρονικό ορίζοντα να εμφανίζει αυτοσυσχέτιση. Μια τέτοια μεταβλητή παρουσιάζουμε στην εργασία αυτή από το χώρο της Οικονομικής του Περιβάλλοντος που είναι «η πυκνότητα των εκπομπών διοξειδίου του άνθρακα ( $CO_2$  σε κιλά ανά ισοδύναμο κιλό

πετρελαίου) στην Ελλάδα»<sup>1</sup>. Έχοντας διαθέσιμες τις τιμές της μεταβλητής αυτής για τα έτη 1961-2010, η εφαρμογή κατάλληλων στατιστικών ελέγχων μας οδηγεί στον ισχυρισμό ότι ο στοχαστικός νόμος γέννησης των τιμών της μεταβλητής αυτής είναι ή το στάσιμο AR(1) ή το στάσιμο ARMA(1,1). Με βάση τα ευρήματα αυτά, στην εργασία αυτή πραγματευόμαστε για πρώτη φορά θέματα εκτιμητικής των εκατοστημορίων όταν ο στοχαστικός νόμος γέννησης των τιμών της μεταβλητής είναι η γενική γραμμική στοχαστική ανέλιξη, ειδικές περιπτώσεις της οποίας αποτελούν τα στάσιμα AR(1) και ARMA(1,1).

Πιο συγκεκριμένα, για την εκτίμηση του  $100 \times P_{th}$  εκατοστημορίου χρησιμοποιούμε τον εκτιμητή μεγίστης πιθανοφάνειας (Dyer *et al.*, 1977) ο οποίος αποτελεί γραμμική συνάρτηση του δειγματικού μέσου και του εκτιμητή μεγίστης πιθανοφάνειας της δειγματικής διακύμανσης. Για τον εκτιμητή μεγίστης πιθανοφάνειας (ML) του  $100 \times P_{th}$  εκατοστημορίου εξάγουμε την ασυμπτωτική του κατανομή από την οποία λαμβάνουμε το αντίστοιχο ασυμπτωτικό διάστημα εμπιστοσύνης. Στη συνέχεια, για την εξέταση της εγκυρότητας της εκτίμησης του διαστήματος αυτού σε πεπερασμένα δείγματα, χρησιμοποιούμε στάσιμες σειρές από το AR(1) και ARMA(1,1) τις οποίες παράγουμε μέσω προσομοιώσεων Monte-Carlo.

Χρησιμοποιώντας τις δημιουργημένες αυτές σειρές, η μελέτη της εγκυρότητας διενεργείται υπολογίζοντας για διαφορετικούς συνδυασμούς μεγέθους δείγματος και αθροιστικής πιθανότητας του εκατοστημορίου τις τιμές δυο στατιστικών κριτηρίων: (α) του πραγματικού επιπέδου εμπιστοσύνης, και (β) του αναμενόμενου ημι-πλάτους του διαστήματος διαρούμενου με το πραγματικό μέγεθος του εκατοστημορίου. Από τις τιμές των δύο αυτών στατιστικών κριτηρίων διαπιστώνουμε ότι η εγκυρότητα του εκτιμηθέντος ασυμπτωτικού διαστήματος σε πεπερασμένα δείγματα εξαρτάται από το μέγεθος του

---

<sup>1</sup> Πηγή των δεδομένων η World Bank (<http://data.worldbank.org/>)

δείγματος, από τη μορφή του στοχαστικού υποδείγματος, AR(1) ή ARMA(1,1), που γεννά τις τιμές της μεταβλητής όταν τα δύο αυτά υποδείγματα έχουν τον ίδιο συντελεστή αυτοσυσχέτισης 1<sup>ης</sup> τάξης, και από την αθροιστική πιθανότητα του εκατοστημορίου.

Η εργασία αυτή κλείνει με εφαρμογή της μεθοδολογίας εκτίμησης των ασυμπτωτικών διαστημάτων εμπιστοσύνης για εκατοστημόρια στην διαθέσιμη σειρά από το χώρο της Οικονομικής του Περιβάλλοντος. Ειδικότερα, εκτιμώντας τα διαστήματα αυτά για εκατοστημόρια που αναφέρονται στην πυκνότητα των εκπομπών διοξειδίου του άνθρακα στην Ελλάδα την περίοδο 1961-2010 για διαφορετικές τιμές της αθροιστικής πιθανότητας P, σχολιάζουμε την εγκυρότητα αυτών βάσει των ευρημάτων που αποκτήθηκαν από τη διεξαγωγή των προσομοιώσεων Monte-Carlo αναφορικά με τις υπολογισθείσες τιμές των δυο παραπάνω στατιστικών κριτηρίων αξιολόγησης.

Με βάση τα παραπάνω, το υπόλοιπο της εργασίας αυτής δομείται ως εξής. Στο επόμενο τμήμα εξάγουμε την ασυμπτωτική κατανομή του εκτιμήτη ML του εκατοστημορίου και δίνουμε τη γενική μορφή του αντίστοιχου ασυμπτωτικού διαστήματος εμπιστοσύνης. Εξειδικεύσεις της γενικής αυτής μορφής εξάγονται για τα στάσιμα υποδείγματα AR(1) και ARMA(1,1). Στο τρίτο τμήμα αιτιολογούμε μέσω εφαρμογής κατάλληλων στατιστικών ελέγχων ότι η σειρά που αναφέρεται στην πυκνότητα των εκπομπών Διοξειδίου του Άνθρακα στην Ελλάδα την περίοδο 1961-2010 έχει ως στοχαστικό νόμο παραγωγής των τιμών της ή το στάσιμο AR(1) ή το στάσιμο ARMA(1,1). Έχοντας διαθέσιμες τις εξειδικευμένες μορφές των ασυμπτωτικών διαστημάτων εμπιστοσύνης εκατοστημορίων για τα δύο αυτά στάσιμα υποδείγματα, στο ίδιο Τμήμα εκτιμούμε τα ασυμπτωτικά διαστήματα για διαφορετικές τιμές της αθροιστικής πιθανότητας του εκατοστημορίου. Ο έλεγχος της εγκυρότητας των εκτιμηθέντων αυτών διαστημάτων διενεργείται μέσω προσομοιώσεων Monte-Carlo στο τέταρτο Τμήμα. Τέλος στο πέμπτο και τελευταίο Τμήμα της εργασίας συνοψίζουμε τα πιο σημαντικά ευρήματα της μελέτης αυτής.



## 2. ΑΣΥΜΠΤΩΤΙΚΑ ΔΙΑΣΤΗΜΑΤΑ

Έστω ότι η διαχρονική εξέλιξη της υπό μελέτη μεταβλητής περιγράφεται από τη γενική γραμμική ανέλιξη,

$$X_t = \mu + \sum_{s=0}^{\infty} \psi_s \varepsilon_{t-s}, \quad (1)$$

όπου  $\sum_{s=0}^{\infty} |\psi_j| < \infty$ , και  $\{\varepsilon_t\}$  ανεξάρτητες τυχαίες μεταβλητές κατανεμόμενες ως  $N(0, \sigma^2)$ .

Με την οριακή κατανομή της  $\{X_t\}$  να είναι η  $N(0, \gamma_o)$ , το  $100 \times P_{th}$  εκατοστημόριο της θα δίνεται από τη σχέση  $K_p = \mu + Z_p \sqrt{\gamma_o}$ , όπου  $Z_p$  είναι η τιμή της αντίστροφης συνάρτησης αθροιστικής κατανομής της  $Z \sim N(0,1)$  υπολογιζόμενη στην αθροιστική πιθανότητα  $P$ .

Για την εκτίμηση του εκατοστημορίου  $K_p$  θα χρησιμοποιηθεί ο «μεροληπτικός» εκτιμητής μεγίστης πιθανοφάνειας  $\hat{K}_p = \bar{X} + Z_p \sqrt{\hat{\gamma}_o}$ , όπου  $\bar{X} = \sum_{t=1}^n X_t / n$  και  $\hat{\gamma}_o = \sum_{t=1}^n (X_t - \bar{X})^2 / n$ . Τα αποτελέσματα των παρακάτω δυο λημμάτων είναι απαραίτητα στην ανάλυση που θα ακολουθήσει για την εξαγωγή του ασυμπτωτικού διαστήματος εμπιστοσύνης για την πραγματική τιμή  $K_p$ .

**Λήμμα 1:** Εάν  $X_t = \mu + \sum_{s=0}^{\infty} \psi_s \varepsilon_{t-s}$ , όπου  $\sum_{s=0}^{\infty} |\psi_j| < \infty$ , με  $\{\varepsilon_t\}$  να είναι ανεξάρτητες τυχαίες μεταβλητές κατανεμόμενες ως  $N(0, \sigma^2)$ , τότε:

(α)  $\sqrt{n}(\bar{X} - \mu)$  ασυμπτωτικά ακολουθεί την κανονική κατανομή με μέσο 0 και

διακύμανση  $\gamma_o \sum_{s=-\infty}^{+\infty} \rho_s$ , και

(β)  $\sqrt{n}(\hat{\gamma}_o - \gamma_o)$  ασυμπτωτικά ακολουθεί την κανονική κατανομή με μέσο 0 και

διακυμάνση  $2\gamma_o^2 \sum_{s=-\infty}^{+\infty} \rho_s^2$ ,

όπου  $\rho_s$  είναι ο συντελεστής αυτοσυσχέτισης  $s^{\text{th}}$  τάξης.

**Απόδειξη:** Βλέπε στο Παράρτημα.

**Λήμμα 2:** Εάν  $X_t = \mu + \sum_{s=0}^{\infty} \psi_s \varepsilon_{t-s}$ , όπου  $\sum_{s=0}^{\infty} |\psi_s| < \infty$ , με  $\{\varepsilon_t\}$  να είναι ανεξάρτητες τυχαίες μεταβλητές κατανεμόμενες ως  $N(0, \sigma^2)$ , τότε για οποιοδήποτε μέγεθος δείγματος η συνδιακύμανση των  $\bar{X}_n$  and  $\hat{\gamma}_0$  είναι μηδέν.

**Απόδειξη:** Βλέπε απόδειξη Πρότασης 2 στο Παράρτημα των Halkos and Kevok (2013b).

Βάσει των αποτελεσμάτων των δυο παραπάνω λημμάτων, το διάνυσμα  $\sqrt{n}[\bar{X} - \mu \quad \hat{\gamma}_0 - \gamma_0]$  ασυμπτωτικά ακολουθεί τη διμεταβλητή κατανομή με μέσο  $\mathbf{0}$  και μήτρα

διακύμανσης-συνδιακύμανσης  $\Sigma = \begin{bmatrix} \gamma_0 \sum_{s=-\infty}^{+\infty} \rho_s & 0 \\ 0 & 2\gamma_0^2 \sum_{s=-\infty}^{+\infty} \rho_s^2 \end{bmatrix}$ . Δοθέντος του P, ισχύει επίσης ότι

$\text{plim} \hat{K}_p = \text{plim} \bar{X}_n + Z_p (\text{plim} \hat{\gamma}_0)^{0.5} = \mu + Z_p \sqrt{\gamma_0} = K_p$ . Επομένως η εφαρμογή της μεθόδου Δέλτα (Knight, 2000, σελ. 149) οδηγεί στην ασυμπτωτική κατανομή του στατιστικού  $\sqrt{n}(\hat{K}_p - K_p)$ , η οποία είναι η κανονική με μέσο μηδέν και διακύμανση  $\mathbf{L}' \cdot \Sigma \cdot \mathbf{L}$  όπου

$$\mathbf{L}' = \begin{bmatrix} \frac{\partial \hat{K}_p}{\partial \bar{X}_n} \Big|_{\substack{\bar{X}_n = \mu \\ \hat{\gamma}_0 = \gamma_0}} & \frac{\partial \hat{K}_p}{\partial \hat{\gamma}_0} \Big|_{\substack{\bar{X}_n = \mu \\ \hat{\gamma}_0 = \gamma_0}} \end{bmatrix} = \begin{bmatrix} 1 & \frac{Z_p}{2\sqrt{\gamma_0}} \end{bmatrix}$$

και  $\mathbf{L}' \cdot \Sigma \cdot \mathbf{L} = \gamma_0 \left\{ \sum_{s=-\infty}^{+\infty} \rho_s + \frac{Z_p^2}{2} \sum_{k=-\infty}^{+\infty} \rho_s^2 \right\}$ . Επομένως το ασυμπτωτικό  $(1-\alpha)100\%$  διάστημα

εμπιστοσύνης για το  $100 \times P_{th}$  εκατοστημόριο δίνεται από τον τύπο

$$\hat{K}_P \pm z_{\alpha/2} \sqrt{\frac{\gamma_o}{n} \left( \sum_{s=-\infty}^{+\infty} \rho_s + \frac{Z_P^2}{2} \sum_{k=-\infty}^{+\infty} \rho_s^2 \right)}. \quad (2)$$

Το διάστημα εμπιστοσύνης της (2) μπορεί να εξειδικευθεί για εναλλακτικά στάσιμα ARMA υποδείγματα. Παρακάτω δίνουμε δύο χαρακτηριστικά παραδείγματα.

### Παράδειγμα 1: Το στάσιμο αυτοπαλίνδρο σχήμα 1<sup>ου</sup> Βαθμού, AR(1).

Το υπόδειγμα αυτό έχει τη γενική μορφή  $X_t = \mu + \phi(X_{t-1} - \mu) + \varepsilon_t$ , με  $|\phi| < 1$ ,  $\gamma_o = \sigma^2 / (1 - \phi^2)$ , και  $\rho_k = \phi^k$  ( $k=0, 1, 2, \dots$ ). Θεωρώντας ότι η ανάλυση έχει ξεκινήσει στο μακρινό παρελθόν, και αντικαθιστώντας διαδοχικά για  $Y_{t-1}$ ,  $Y_{t-2}$ ,  $Y_{t-3}$ , ..., το AR(1) λαμβάνει τη μορφή της γενικής γραμμικής ανάλυξης της σχέσης (1) με  $\psi_j = \phi^j$ . Επιπλέον ισχύει ότι

$$\sum_{s=-\infty}^{\infty} \rho_s = 1 + \frac{2\phi}{1-\phi} = \frac{1+\phi}{1-\phi} \quad \text{και} \quad \sum_{s=-\infty}^{\infty} \rho_s^2 = 1 + \frac{2\phi^2}{1-\phi^2} = \frac{1+\phi^2}{1-\phi^2}.$$

Επομένως το διάστημα εμπιστοσύνης στην (2) για το AR(1) εξειδικεύεται ως

$$\hat{K}_P \pm z_{\alpha/2} \sqrt{\frac{\gamma_o}{n} \left( \frac{1+\rho_1}{1-\rho_1} + \frac{Z_R^2}{2} \frac{1+\rho_1^2}{1-\rho_1^2} \right)}, \quad (3)$$

καθώς ισχύει  $\rho_1 = \phi$ .

### Παράδειγμα 2: Το στάσιμο και αντιστρέψιμο σχήμα ARMA(1,1).

Το υπόδειγμα αυτό έχει τη γενική μορφή  $X_t = \mu + \phi(X_{t-1} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1}$ , με  $|\phi| < 1$ ,  $|\theta| < 1$ ,  $\gamma_o = \frac{1+\theta^2+2\phi\theta}{1-\phi^2} \sigma^2$ ,  $\rho_1 = \frac{(1+\phi\theta)(\phi+\theta)}{1+\theta^2+2\phi\theta}$ , και  $\rho_k = \phi^{k-1} \rho_1$  για  $k \geq 2$ . Θεωρώντας ότι η ανάλυση έχει ξεκινήσει στο μακρινό παρελθόν, ο Harvey (1993, σελ. 26) δείχνει ότι το ARMA(1,1) λαμβάνει τη μορφή της γενικής γραμμικής ανάλυξης της σχέσης (1) με  $\psi_o = 1$ ,  $\psi_1 = \phi + \theta$ , και  $\psi_k = \phi \psi_{k-1}$  για  $k \geq 2$ . Επιπλέον ισχύουν οι παρακάτω σχέσεις

$$\sum_{s=-\infty}^{\infty} \rho_s = 1 + 2\rho_1/(1-\phi) \quad \text{και} \quad \sum_{s=-\infty}^{\infty} \rho_s^2 = 1 + 2\rho_1^2/(1-\phi^2).$$

Επομένως το διάστημα εμπιστοσύνης στην (1) για το ARMA(1,1) εξειδικεύεται ως

$$\hat{K}_P \pm Z_{\alpha/2} \sqrt{\frac{\gamma_o}{n} \left( 1 + \frac{2\rho_1^2}{\rho_1 - \rho_2} + \frac{Z_R^2}{2} \left( 1 + \frac{2\rho_1^4}{\rho_1^2 - \rho_2^2} \right) \right)}, \quad (4)$$

μετά την αντικατάσταση του  $\phi$  από το λόγο  $\rho_2/\rho_1$ .

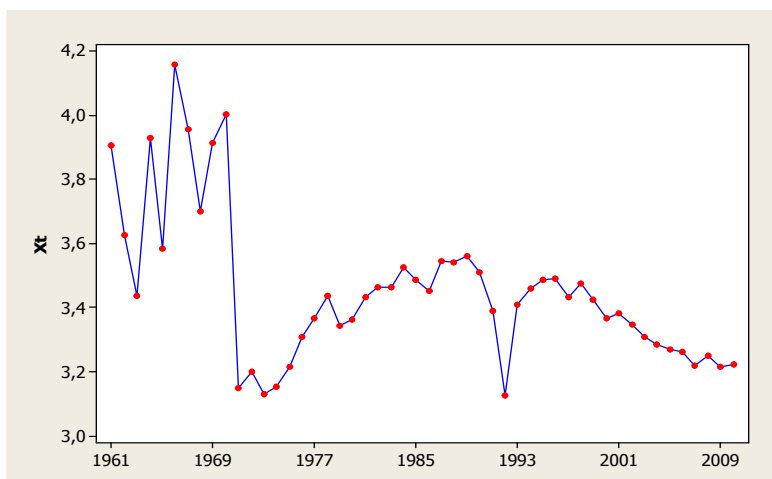
### 3. ΕΦΑΡΜΟΓΗ ΣΕ ΠΕΡΙΒΑΛΛΟΝΤΙΚΑ ΔΕΔΟΜΕΝΑ

Για τα έτη 1961-2010, το Διάγραμμα 1 παρουσιάζει τη διαχρονική εξέλιξη της υπό μελέτη περιβαλλοντικής μεταβλητής,  $X_t$ , που είναι η πυκνότητα των εκπομπών Διοξειδίου του Άνθρακα ( $\text{CO}_2$ , σε κιλά ανά ισοδύναμο κιλό πετρελαίου) στην Ελλάδα. Για τον έλεγχο της στασιμότητας της σειράς, εφαρμόστηκαν έλεγχοι ADF, τα αποτελέσματα των οποίων παρουσιάζονται στον Πίνακα 1. Έχοντας ως ανεξάρτητη μεταβλητή την  $\Delta X_t$  και ως ερμηνευτικές, εκτός του σταθερού όρου, τη μεταβλητή  $X_{t-1}$ , τη μεταβλητή χρονικής τάσης,  $t$ , και τις μεταβλητές χρονικών υστερήσεων διαφορών  $\Delta X_{t-j}$ , εναλλακτικές εξισώσεις εκτιμήθηκαν για  $j=1, \dots, 10$ .

Η χρήση του κριτηρίου Schwarz έδειξε ότι στην εκτιμηθείσα εξίσωση δεν πρέπει να χρησιμοποιηθούν μεταβλητές χρονικών υστερήσεων διαφορών. Χωρίς τη χρήση των μεταβλητών αυτών, η εκτιμηθείσα εξίσωση παρουσιάζεται στον Πίνακα 1 ως εκτιμηθέν Υπόδειγμα 1. Παρατηρούμε όμως, ότι στο υπόδειγμα αυτό ο συντελεστής της μεταβλητής χρονικής τάσης είναι μη στατιστικά σημαντικός σε επίπεδο 5%. Χωρίς τη μεταβλητή χρονικής τάσης, η νέα εκτιμηθείσα εξίσωση παρουσιάζεται επίσης στο Πίνακα 1 ως εκτιμηθέν υπόδειγμα 2. Παρόλα αυτά, χρησιμοποιώντας είτε το εκτιμηθέν υπόδειγμα 1 είτε το εκτιμηθέν υπόδειγμα 2, παρατηρούμε ότι σε επίπεδο στατιστικής σημαντικότητας 5%, η

υπόθεση της μοναδιαίας ρίζας απορρίπτεται, γεγονός το οποίο μας οδηγεί να ισχυριστούμε ότι η υπό μελέτη σειρά είναι στάσιμη.

**Διάγραμμα 1:** Διαχρονική εξέλιξη της πυκνότητας εκπομπών διοξειδίου του Άνθρακα (CO<sub>2</sub>) (σε κιλά ανά ισοδύναμο κιλό πετρελαίου) στην Ελλάδα το διάστημα 1961-2010.



**Πίνακας 1:** Αποτελέσματα στατιστικών ελέγχων Augmented Dickey Fuller (ADF)

	Constant term	Time-Trend	X <sub>t-1</sub>	ADF Test Value	ADF p-value
<b>Εκτιμηθέν Υπόδειγμα 1</b>					
Coefficient:	1.806002	-0.003701	-0.499376	-3.996764	0.0152
p-value:	0.0003	0.0815	0.0002		
<b>Εκτιμηθέν Υπόδειγμα 2</b>					
Coefficient:	1.318230		-0.385110	-3.511855	0.0117
p-value:	0.0010		0.0011		

Στα Διαγράμματα 2 και 3 παρουσιάζουμε αντίστοιχα τη συνάρτηση δειγματικής αυτοσυσχέτισης και τη συνάρτηση δειγματικής μερικής αυτοσυσχέτισης. Οι μορφές των δειγματικών αυτών συναρτήσεων είναι ενδεικτικές είτε για υπόδειγμα AR(1) είτε για ARMA(1,1). Με τη χρήση της μεθόδου OLS παρουσιάζουμε παρακάτω τα εκτιμημένα δυο αυτά υποδείγματα με τα p-values να δίνονται στις παρενθέσεις:

$$\text{Εκτιμημένο AR(1): } \hat{Y}_t = 1.318230 + 0.614890 Y_{t-1} + \hat{\varepsilon}_t$$

$$(0.0011) \quad (0.0000)$$

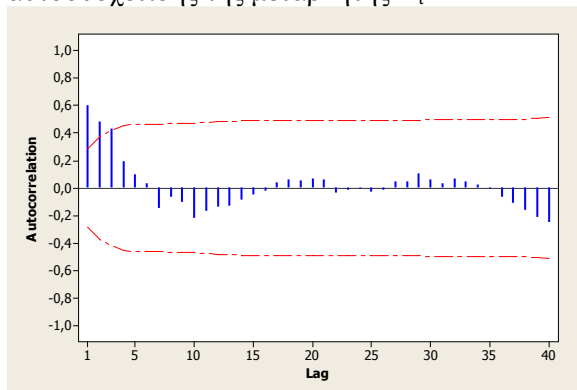
$$\hat{\sigma}_{\varepsilon} = 0.180552,$$

$$\text{Εκτιμημένο ARMA(1): } \hat{Y}_t = 3.402104 + 0.782531 Y_{t-1} - 0.285921 \hat{\varepsilon}_{t-1} + \hat{\varepsilon}_t$$

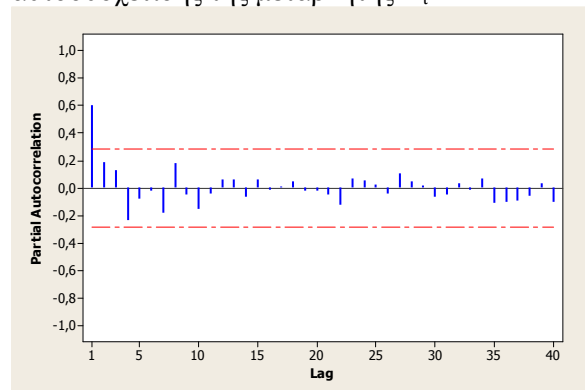
$$(0.0000) \quad (0.0000) \quad (0.1706)$$

$$\hat{\sigma}_{\varepsilon} = 0.179286.$$

**Διάγραμμα 2:** Συνάρτηση δειγματικής αυτοσυσχέτισης της μεταβλητής  $X_t$



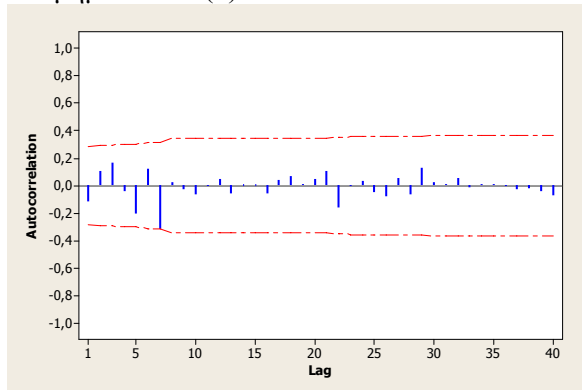
**Διάγραμμα 3** Συνάρτηση δειγματικής μερικής αυτοσυσχέτισης της μεταβλητής  $X_t$



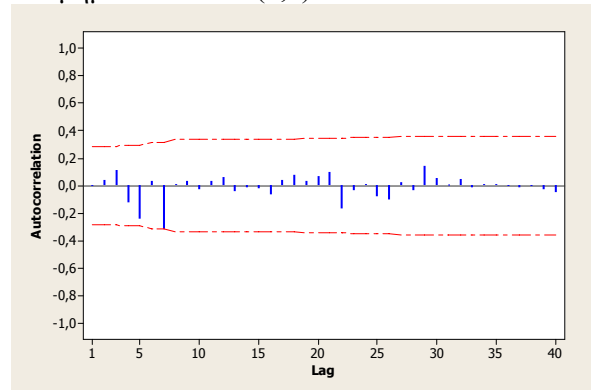
Για την περαιτέρω αξιολόγηση των εκτιμηθέντων  $AR(1)$  και  $ARMA(1,1)$ , στα διαγράμματα 4 και 5 παρουσιάζουμε τις συναρτήσεις αυτοσυσχέτισης των καταλοίπων για τα δύο υποδείγματα αντίστοιχα. Επιπλέον στον Πίνακα 2 δίνουμε τα αποτελέσματα και τις τιμές  $P$  ( $p$ -values) των διαγνωστικών ελέγχων Jarque-Bera για κανονικότητα και ARCH-LM για αυτοπαλίνδρομη υπό συνθήκη ετεροσκεδαστικότητα. Καταρχάς παρατηρούμε στα δύο διαγράμματα ότι κανένας δειγματικός συντελεστής αυτοσυσχέτισης των καταλοίπων δεν βρίσκεται εκτός των δυο γραμμών οριοθέτησης. Άρα τα σφάλματα και στα δυο υποδείγματα εμφανίζονται ως ασυσχέτιστα. Επιπλέον τα  $p$ -values των στατιστικών  $F$  και  $nR^2$  δίνουν ισχυρές ενδείξεις στο να αποφανθούμε ότι στα σφάλματα των δυο υποδειγμάτων δεν υπάρχει αποτέλεσμα ARCH.

Το πρόβλημα που ανακύπτει στα κατάλοιπα των εκτιμηθέντων δύο υποδειγμάτων αφορά τις τιμές  $P$  του ελέγχου Jarque-Bera, τα οποία υποδεικνύουν ότι τα σφάλματα δεν ακολουθούν τη κανονική κατανομή. Παρόλα αυτά, εμείς θα παραβλέψουμε το πρόβλημα αυτό και θα συνεχίσουμε στην εκτίμηση των εκατοστημορίων και των αντίστοιχων ασυμπτωτικών διαστημάτων εμπιστοσύνης για την υπό εξέταση μεταβλητή όταν ο στοχαστικός νόμος που παράγει τις τιμές αυτής είναι είτε το  $AR(1)$  είτε το  $ARMA(1,1)$ .

**Διάγραμμα 4:** Συνάρτηση δειγματικής αυτοσυσχέτισης των καταλοίπων του εκτιμημένου AR(1)



**Διάγραμμα 5:** Συνάρτηση δειγματικής αυτοσυσχέτισης των καταλοίπων του εκτιμημένου ARMA(1,1)



**Πίνακας 2:** Αποτελέσματα διαγνωστικών ελέγχων για τα κατάλοιπα των εκτιμηθέντων AR(1) και ARMA(1,1)

Test	Jarque-Bera	ARCH <sub>LM</sub> (1)	
		F-statistic	n*R2
AR(1) Test statistic p-value	65.93892 (0.0000)	0.083217 (0.7743)	0.086678 (0.7684)
ARMA(1,1) Test statistic p-value	49.63137 (0.0000)	0.415192 (0.5225)	0.429369 (0.5123)

Για διαφορετικές τιμές της αθροιστικής πιθανότητας P του εκατοστημορίου, στον Πίνακα 3 δίνουμε για τα δύο υποδείγματα τις εκτιμήσεις των εκατοστημορίων και τα κάτω και πάνω όρια των 95% ασυμπτωτικών διαστημάτων εμπιστοσύνης, οι μορφές των οποίων δίνονται στις σχέσεις (3) και (4). Για την εκτίμηση του δειγματοληπτικού σφάλματος στις δύο αυτές μορφές των ασυμπτωτικών διαστημάτων εμπιστοσύνης, το  $\gamma_0$  αντικαταστάθηκε από την εκτίμησή του  $\hat{\gamma}_0$ , ενώ οι συντελεστές αυτοσυσχέτισης  $\hat{\rho}_1$  και  $\hat{\rho}_2$  εκτιμήθηκαν βάσει του τύπου (Harvey, 1993, σελ. 11)

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^n (X_t - \bar{X})(X_{t-s} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}, \quad s=1,2.$$

**Πίνακας 3:** Εκτιμήσεις για τα εκατοστημόρια της πυκνότητας των εκπομπών διοξειδίου του άνθρακα (CO<sub>2</sub>, σε κιλά ανά ισοδύναμο κιλό πετρελαίου) στην Ελλάδα. [P η αθροιστική πιθανότητα του εκατοστημορίου,  $\hat{K}_p$  η εκτίμηση του εκατοστημορίου, LL και UL τα πάνω και κάτω όρια του 95% διαστήματος εμπιστοσύνης για το  $K_p$ , και REHL το ήμισυ της διαφοράς (UL–LL) διαιρούμενης με  $\hat{K}_p$ ].

P	$\hat{K}_p$	AR(1)			ARMA(1,1)		
		LL	UL	REHL	LL	UL	REHL
0.2	3.256484	3.114648	3.398319	0.0436	3.069941	3.443027	0.0573
0.3	3.331078	3.196283	3.465873	0.0405	3.152175	3.509981	0.0537
0.4	3.394817	3.263583	3.526050	0.0387	3.219751	3.569882	0.0516
0.45	3.424842	3.294440	3.555244	0.0381	3.250669	3.599015	0.0509
0.55	3.483941	3.353539	3.614343	0.0374	3.309768	3.658114	0.0500
0.5	3.454391	3.324262	3.584521	0.0377	3.280511	3.628272	0.0503
0.6	3.513966	3.382733	3.645200	0.0373	3.338900	3.689032	0.0498
0.8	3.652299	3.510464	3.794135	0.0388	3.465756	3.838842	0.0511
0.9	3.755749	3.599815	3.911683	0.0415	3.553718	3.957780	0.0538
0.95	3.841180	3.670611	4.011749	0.0444	3.622853	4.059507	0.0568
0.99	4.001434	3.798314	4.204554	0.0508	3.746279	4.256589	0.0638
0.999	4.181062	3.936410	4.425713	0.0585	3.878137	4.483987	0.0725

Από την εξέταση των στοιχείων του Πίνακα 3 διαπιστώνουμε τα εξής: (α) Όπως αναμενόταν η αύξηση της αθροιστικής πιθανότητας P οδηγεί σε μεγαλύτερες εκτιμήσεις για το εκατοστημόριο  $K_p$ , και (β) για τα δύο υποδείγματα, η σχετική ακρίβεια των διαστημάτων εμπιστοσύνης μειώνεται (δηλαδή το πλάτος του διαστήματος εμπιστοσύνης ως ποσοστό του  $K_p$  αυξάνεται) με την αύξηση του P.

Κλείνοντας το τμήμα αυτό, είναι αξιοσημείωτο να θέσουμε δυο ζητήματα. Το πρώτο αφορά τη σύγκριση των διαστημάτων εμπιστοσύνης μεταξύ του AR(1) και του ARMA(1,1). Μια τέτοια σύγκριση στο δείγμα των 50 ετών δεν είναι εφικτή για δυο λόγους. Ο πρώτος είναι ότι ο συντελεστής του  $\hat{\varepsilon}_{t-1}$  στο εκτιμηθέν ARMA(1,1) δεν είναι όπως φαίνεται στατιστικά σημαντικός βάσει μιας τιμής P την οποία δεν γνωρίζουμε σε ποια τιμή θα διαμορφωνόταν (μικρή ή μεγάλη) εάν είχαμε ένα δείγμα αρκούντως μεγάλο. Ο δεύτερος λόγος είναι ο διαφορετικός εκτιμημένος συντελεστής αυτοσυσχέτισης πρώτης τάξης για τα δύο υποδείγματα. Πιο συγκεκριμένα, χρησιμοποιώντας τις εκτιμηθείσες τιμές των  $\phi$  και  $\theta$  στον αντίστοιχο τύπο του θεωρητικού συντελεστή αυτοσυσχέτισης πρώτης τάξης του κάθε



υποδείγματος θα παίρναμε για το AR(1)  $\hat{\rho}_1 = 0.614890$  και για το ARMA(1,1)  $\hat{\rho}_1 = 0.607784$ .

Το δεύτερο ζήτημα αφορά την ερμηνεία του κάθε διαστήματος εμπιστοσύνης για το διαφορετικό εκατοστημόριο  $K_p$ . Η αθροιστική πιθανότητα  $P$  του εκατοστημορίου αφορά για τη συγκεκριμένη σειρά ποσοστό ετών. Επομένως για  $P=0.80$ , μπορούμε να ισχυριστούμε ότι εάν οι εξωτερικές συνθήκες διαμόρφωσης των τιμών της σειράς παραμείνουν οι ίδιες, στο 80% των ετών, με πιθανότητα 95% η πυκνότητα των εκπομπών διοξειδίου του άνθρακα θα κυμανθεί μεταξύ 3.510464 και 3.794135 σε κιλά ανά ισοδύναμο κιλό πετρελαίου στην περίπτωση του AR(1) και μεταξύ 3.465756 και 3.838842 κιλών για το ARMA(1,1).

#### **4. ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΕΓΚΥΡΟΤΗΤΑΣ ΤΩΝ ΑΣΥΜΠΤΩΤΙΚΩΝ ΔΙΑΣΤΗΜΑΤΩΝ ΕΜΠΙΣΤΟΣΥΝΗΣ**

Στο τμήμα αυτό ελέγχουμε την εγκυρότητα των κατασκευασθέντων διαστημάτων εμπιστοσύνης για τις πραγματικές τιμές των εκατοστημορίων της οριακής κατανομής των εκπομπών διοξειδίου του άνθρακα, όταν ο στοχαστικός νόμος που παράγει τις τιμές της συγκεκριμένης μεταβλητής περιγράφεται είτε από το στάσιμο AR(1) είτε από το στάσιμο ARMA(1,1). Η διερεύνηση αυτή διενεργείται υπολογίζοντας τα παρακάτω δυο στατιστικά κριτήρια σε 20000 σειρές στάσιμων AR(1) και σε 20000 σειρές στάσιμων ARMA(1,1) οι οποίες παρήχθησαν μέσω προσωμοιώσεων Monte-Carlo:

(α) **Κάλυψη (Coverage)** που δείχνει το πραγματικό επίπεδο εμπιστοσύνης που επιτυγχάνει σε πεπερασμένα δείγματα το εκτιμημένο ασυμπτωτικό διάστημα εμπιστοσύνης, και

(β) **Σχετικό Αναμενόμενο Ημι-πλάτος** του διαστήματος (Relative Expected Half-Length, REHL).

Ας σημειωθεί επίσης ότι η γεννήτρια τυχαίων αριθμών που χρησιμοποιήθηκε καθώς και η μέθοδος παραγωγής τυχαίων τιμών από τη τυποποιημένη κανονική κατανομή περιγράφονται στον Kevork (2010).

Η επιλογή των παραμέτρων για τα υποδείγματα AR(1) και ARMA(1,1) έγινε βάσει δυο αρχών: (α) Η οριακή κατανομή της υπό μελέτη μεταβλητής να έχει τον ίδιο στάσιμο μέσο  $\mu = 100$  και την ίδια διακύμανση  $\gamma_0 = 400$ , και (β) ο συντελεστής αυτοσυσχέτισης 1<sup>ης</sup> τάξης να έχει την ίδια τιμή και στα δυο υποδείγματα. Με βάση τις δυο αυτές αρχές, οι τιμές των παραμέτρων για τα δυο υποδείγματα είναι οι εξής:

$\rho_1 = 0.56$	
AR(1):	$\phi = 0.56, \quad \sigma_\varepsilon^2 = 274.56$
ARMA(1,1):	$\phi = 0.3, \quad \theta = 0.4, \quad \sigma_\varepsilon^2 = 260$

Για την επίτευξη στασιμότητας σε κάθε μια από τις 20000 προσομοιωμένες σειρές από το κάθε υπόδειγμα, η τιμή  $X_0$  παρήχθη από τη στάσιμη οριακή κατανομή  $N(100,400)$ , και επιπλέον για το ARMA(1,1) η αρχική τιμή  $\varepsilon_0$  παρήχθη από την κατανομή των σφαλμάτων  $N(0,260)$ . Στη συνέχεια, η διερεύνηση της εγκυρότητας των εκτιμηθέντων ασυμπτωτικών διαστημάτων εμπιστοσύνης, όπως αυτά δίνονται στις σχέσεις (3) και (4), διενεργήθηκε για διαφορετικούς συνδυασμούς μεγέθους δείγματος και διαφορετικές τιμές της αθροιστικής πιθανότητας,  $P$ , του εκατοστημορίου. Ειδικότερα, για κάθε μια από τις 20000 παραχθείσες σειρές από το κάθε υπόδειγμα, και για κάθε ένα συνδυασμό μεγέθους δείγματος,  $n$ , και αθροιστικής πιθανότητας,  $P$ , εκτιμήθηκαν το εκατοστημόριο, το δειγματοληπτικό σφάλμα του διαστήματος εμπιστοσύνης, και τα όρια του διαστήματος εμπιστοσύνης.

Με βάση τις εκτιμήσεις αυτές, για κάθε ένα συνδυασμό  $n$  και  $P$ , υπολογίσθηκαν η κάλυψη ως το ποσοστό των 20000 διαστημάτων εμπιστοσύνης τα οποία περιείχαν την πραγματική τιμή του εκατοστημορίου, και το REHL ως ο μέσος αριθμητικός του ημι-πλάτους των 20000 διαστημάτων εμπιστοσύνης διαιρούμενος με την πραγματική τιμή του εκατοστημορίου. Η διαίρεση του αναμενόμενου Ημι-πλάτους του διαστήματος με την πραγματική τιμή του εκατοστημορίου κρίνεται αναγκαία για τη διασφάλιση της

συγκρισιμότητας της ακρίβειας διαστημάτων εμπιστοσύνης που κατασκευάζονται για διαφορετικές τιμές του  $P$ . Και αυτό γιατί με την αύξηση του  $P$ , αυξάνεται και η τιμή του εκατοστημορίου.

Ένα πρόβλημα που ανέκυψε σε ορισμένες σειρές από το ARMA(1,1) ήταν ότι σε χαμηλά μεγέθη δείγματος η εκτίμηση της διακύμανσης του  $\hat{K}_p$  ήταν αρνητική. Στο Πίνακα 4 δίνουμε τον αριθμό των σειρών αυτών για διαφορετικές τιμές του  $P$ . Παρατηρούμε ότι ένας σημαντικός αριθμός τέτοιων σειρών εμφανίζεται για πολύ χαμηλά μεγέθη δείγματος της τάξεως των 5 και 10 παρατηρήσεων, ενώ ο αριθμός αυτός αρχίζει να γίνεται αμελητέος έως και μηδενικός για μεγέθη δείγματος άνω των 30 παρατηρήσεων. Επιπλέον, για τα πολύ χαμηλά μεγέθη δείγματος, ο αριθμός των σειρών αυτών αυξάνει όσο μετακινούμαστε στη τιμή  $P=0.50$  για την οποία λαμβάνουμε το μέσο αριθμητικό.

**Πίνακας 4:** Αριθμός στάσιμων σειρών ARMA(1,1) για τις οποίες η ασυμπτωτική διακύμανση του  $\hat{K}_p$  εκτιμώμενη σε πεπερασμένα δείγμα είναι αρνητική. [Συνολικός αριθμός σειρών για στάσιμο ARMA(1,1) που παρήχθησαν μέσω προσομοιώσεων Monte-Carlo ήταν 20000].

		Μέγεθος Δείγματος								
$\rho_1=0.56$		n=5	n=10	n=20	n=30	n=40	n=50	n=60	n=80	n=100
ARMA(1,1)	P=0.2	650	119	14	3	0	0	0	0	0
	P=0.3	806	100	8	0	0	0	0	0	0
	P=0.4	921	87	3	0	0	0	0	0	0
	P=0.5	997	85	2	0	0	0	0	0	0
	P=0.6	921	87	3	0	0	0	0	0	0
	P=0.7	806	100	8	0	0	0	0	0	0
	P=0.8	650	119	14	3	0	0	0	0	0
	P=0.99	367	178	39	5	1	0	0	0	0

Στη συνέχεια για τα δύο υποδείγματα AR(1) και ARMA(1,1), και για ονομαστικό επίπεδο εμπιστοσύνης 95%, στον Πίνακα 5 δίνουμε τις τιμές της κάλυψης ενώ στον Πίνακα 6 τις τιμές του REHL. Για το ARMA(1,1), σε χαμηλά μεγέθη δείγματος, οι τιμές των δύο αυτών στατιστικών κριτηρίων υπολογίσθηκαν ως μέσοι όροι των αριθμών των σειρών για τις οποίες η εκτιμηθείσα διακύμανση του  $\hat{K}_p$  ήταν θετική. Από τα στοιχεία του Πίνακα 5 διαπιστώνουμε ότι σε χαμηλά μεγέθη δείγματος η κάλυψη είναι σημαντικά μικρότερη του

95%, ενώ με την αύξηση του μεγέθους του δείγματος η ταχύτητα σύγκλισης της κάλυψης στο ονομαστικό επίπεδο εμπιστοσύνης είναι σχετικά πιο γρήγορη στο ARMA(1,1). Δεχόμενοι επίσης ότι μια κάλυψη της τάξεως άνω του 90% αποτελεί μια αποδεκτή προσέγγιση του ονομαστικού επιπέδου εμπιστοσύνης, ένα δείγμα 50 παρατηρήσεων και πάνω διασφαλίζει την εγκυρότητα του ασυμπτωτικού διαστήματος εμπιστοσύνης. Αντίθετα εάν γίνουμε άκρως αυστηροί στην προσέγγιση αυτή και επιθυμούμε καλύψεις πχ. άνω του 94%, τότε το απαραίτητο μέγεθος δείγματος είναι της τάξης των 500 παρατηρήσεων κάτι που από πλευράς δεδομένων της πραγματικής οικονομίας είναι αδύνατον να είναι διαθέσιμο.

**Πίνακας 5:** Καλύψεις των εκτιμηθέντων ασυμπτωτικών διαστημάτων εμπιστοσύνης για εκατοστημόρια σε 20000 στάσιμες σειρές AR(1) and ARMA(1,1) που παρήχθησαν μέσω προσομοιώσεων Monte-Carlo. Το ονομαστικό επίπεδο εμπιστοσύνης είναι 95%

$\rho_1=0.56$		n=5	n=10	n=20	n=30	n=50	n=100	n=200	n=500	n=1000	n=2000
P=0.5	ARMA	0.64	0.78	0.86	0.89	0.91	0.93	0.94	0.94	0.95	0.95
	AR	0.56	0.73	0.83	0.87	0.90	0.92	0.93	0.94	0.95	0.95
P=0.6	ARMA	0.64	0.78	0.86	0.89	0.91	0.93	0.94	0.94	0.95	0.95
	AR	0.56	0.73	0.83	0.87	0.90	0.92	0.94	0.94	0.95	0.95
P=0.8	ARMA	0.64	0.78	0.86	0.89	0.91	0.92	0.94	0.95	0.95	0.95
	AR	0.59	0.74	0.83	0.87	0.89	0.92	0.94	0.94	0.95	0.95
P=0.95	ARMA	0.62	0.76	0.85	0.88	0.91	0.92	0.94	0.94	0.95	0.95
	AR	0.61	0.73	0.82	0.86	0.89	0.91	0.93	0.94	0.95	0.95
P=0.99	ARMA	0.61	0.74	0.84	0.87	0.90	0.92	0.94	0.94	0.95	0.95
	AR	0.60	0.72	0.81	0.85	0.88	0.91	0.93	0.94	0.95	0.95

Παρατηρώντας επίσης τα στοιχεία του Πίνακα 6 βλέπουμε ότι για κάθε μέγεθος δείγματος, το REHL είναι μικρότερο στο ARMA(1,1) όταν ο θεωρητικός συντελεστής αυτοσυσχέτισης 1<sup>ης</sup> τάξης και των δυο υποδειγμάτων είναι ο ίδιος. Επίσης, για κάθε υπόδειγμα, το REHL βαίνει μειούμενο όσο αυξάνεται το P έως κάποια τιμή του μεταξύ του 0.6 και 0.8, και στη συνέχεια αυξάνεται και πάλι.

Λαμβάνοντας λοιπόν υπόψη τις παραπάνω διαπιστώσεις για τις ιδιότητες των ασυμπτωτικών διαστημάτων εμπιστοσύνης όταν αυτά εκτιμώνται σε πεπερασμένα δείγματα, μπορούμε να ισχυριστούμε ότι τα διαστήματα εμπιστοσύνης για τα εκατοστημόρια της οριακής κατανομής των εκπομπών διοξειδίου του άνθρακα τα οποία δόθηκαν στο

προηγούμενο τμήμα είναι έγκυρα. Καταρχάς, βλέποντας ότι οι εκτιμηθέντες συντελεστές αυτοσυσχέτισης 1<sup>ης</sup> τάξης για τα δύο υποδείγματα είναι σχετικά κοντά στη τιμή  $\rho_1 = 0.56$ , μπορούμε να ισχυριστούμε ότι η απώλεια επι της % από το ονομαστικό επίπεδο εμπιστοσύνης 95% θα κυμαίνεται περίπου στο 5%.

**Πίνακας 6:** REHLs των εκτιμηθέντων ασυμπτωτικών διαστημάτων εμπιστοσύνης για εκατοστημόρια σε 20000 στάσιμες σειρές AR(1) and ARMA(1,1) που παρήχθησαν μέσω προσομοιώσεων Monte-Carlo. Το ονομαστικό επίπεδο εμπιστοσύνης είναι 95%

$\rho_1=0.56$		n=10	n=20	n=30	n=50	n=100	n=200	n=500	n=1000	n=2000
R=0.2	ARMA	0.1855	0.1601	0.1383	0.1116	0.0814	0.0587	0.0374	0.0266	0.0188
	AR	0.1915	0.1716	0.1516	0.1246	0.0923	0.0669	0.0429	0.0305	0.0216
R=0.6	ARMA	0.1316	0.1141	0.0991	0.0802	0.0586	0.0423	0.0270	0.0191	0.0136
	AR	0.1371	0.1246	0.1105	0.0910	0.0674	0.0489	0.0314	0.0223	0.0158
R=0.7	ARMA	0.1300	0.1126	0.0976	0.0789	0.0576	0.0415	0.0265	0.0188	0.0133
	AR	0.1351	0.1221	0.1081	0.0890	0.0659	0.0478	0.0306	0.0218	0.0154
R=0.8	ARMA	0.1321	0.1140	0.0985	0.0795	0.0580	0.0418	0.0266	0.0189	0.0134
	AR	0.1363	0.1221	0.1079	0.0887	0.0657	0.0476	0.0305	0.0217	0.0154
R=0.9	ARMA	0.1393	0.1195	0.1029	0.0828	0.0603	0.0434	0.0277	0.0196	0.0139
	AR	0.1423	0.1259	0.1110	0.0910	0.0673	0.0488	0.0313	0.0222	0.0157
R=0.99	ARMA	0.1647	0.1395	0.1194	0.0957	0.0695	0.0499	0.0318	0.0226	0.0160
	AR	0.1647	0.1428	0.1252	0.1023	0.0755	0.0547	0.0350	0.0249	0.0176

Επιπλέον, για το κάθε υπόδειγμα, οι μεταβολές που παρατηρούνται στο REHL στις προσομοιωμένες σειρές όταν αυξάνεται το P είναι ακριβώς οι ίδιες με αυτές που παρατηρούνται από τα πραγματικά δεδομένα της σειράς των εκπομπών διοξειδίου του άνθρακα. Το ότι το REHL εμφανίζεται μικρότερο στο AR(1), αυτό οφείλεται στο γεγονός ότι στο AR(1) ο εκτιμηθείς συντελεστής αυτοσυσχέτισης 1<sup>ης</sup> τάξης είναι μεγαλύτερος από τον αντίστοιχο συντελεστή στο ARMA(1,1). Τέλος, και στα δυο εκτιμηθέντα υποδείγματα, οι επιπτώσεις της μη κανονικότητας των σφαλμάτων πάνω στην εγκυρότητα των διαστημάτων εμπιστοσύνης θα το παραπέμψουμε ως θέμα μελλοντικής έρευνας.

## 5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην εργασία αυτή ασχοληθήκαμε με θέματα εκτιμητικής εκατοστημορίων όταν η διαχρονική εξέλιξη της υπό μελέτη μεταβλητής καθορίζεται από τη γενική γραμμική ανέλιξη ειδικές περιπτώσεις της οποίας είναι τα στάσιμα υποδείγματα  $AR(1)$  και  $ARMA(1,1)$ . Χρησιμοποιώντας τον εκτιμητή μεγίστης πιθανοφάνειας, εξάγουμε την ασυμπτωτική του κατανομή και δίνουμε τη γενική μορφή του αντίστοιχου ασυμπτωτικού διαστήματος εμπιστοσύνης για την πραγματική τιμή του εκατοστημορίου. Ειδικές μορφές του ασυμπτωτικού αυτού διαστήματος πήραμε για τα στάσιμα  $AR(1)$  και  $ARMA(1,1)$ .

Την εγκυρότητα του ασυμπτωτικού διαστήματος όταν αυτό εκτιμάται σε πεπερασμένα δείγματα ελέγξαμε σε στάσιμες σειρές  $AR(1)$  και  $ARMA(1,1)$  τις οποίες παράγαμε μέσω προσομοιώσεων Monte Carlo. Ας σημειωθεί εδώ ότι οι παράμετροι των δύο υποδειγμάτων επιλέχθηκαν έτσι ώστε η οριακή κατανομή των δυο υποδειγμάτων να έχει τον ίδιο μέσο και την ίδια διακύμανση καθώς και ο θεωρητικός συντελεστής αυτοσυσχέτισης 1<sup>ης</sup> τάξης να είναι ο ίδιος και ίσος με 0.56. Σε ονομαστικό επίπεδο εμπιστοσύνης 95% βρήκαμε ότι με δείγμα τουλάχιστον 50 παρατηρήσεων το εκτιμημένο διάστημα εμπιστοσύνης επιτυγχάνει πραγματικό επίπεδο εμπιστοσύνης τουλάχιστον 90%. Επίσης με την αύξηση της αθροιστικής πιθανότητας που συνδέεται με το εκάστοτε εκατοστημόριο παρατηρήσαμε ότι το αναμενόμενο πλάτος του διαστήματος διαιρούμενο με την πραγματική τιμή του εκατοστημορίου στην αρχή μειώνεται και μετά αυξάνεται, με το ελάχιστο να συμβαίνει όταν η αθροιστική πιθανότητα βρίσκεται μεταξύ του 0.6 και του 0.8.

Έχοντας διαθέσιμη τη σειρά που αναφέρεται στην πυκνότητα των εκπομπών διοξειδίου του άνθρακα ( $CO_2$ , σε κιλά ανά ισοδύναμο κιλό πετρελαίου) στην Ελλάδα από το 1961 έως το 2010 και εφαρμόζοντας κατάλληλους στατιστικούς ελέγχους καταλήξαμε ότι ο στοχαστικός νόμος που έχει παράγει τις τιμές της μεταβλητής αυτής είναι είτε το στάσιμο  $AR(1)$  είτε το στάσιμο  $ARMA(1,1)$ . Στην σειρά αυτή εφαρμόσαμε τη μεθοδολογία εκτίμησης

του ασυμπτωτικού διαστήματος εμπιστοσύνης δίνοντας ξεχωριστά για το AR(1) και ξεχωριστά για το ARMA(1,1) το κάτω και πάνω όριο του εκτιμηθέντος ασυμπτωτικού διαστήματος εμπιστοσύνης για διαφορετικές τιμές της αθροιστικής πιθανότητας του εκατοστημορίου.

Από τα ευρήματα του ελέγχου αξιοπιστίας των εκτιμηθέντων ασυμπτωτικών διαστημάτων εμπιστοσύνης μέσω των πειραμάτων προσομοίωσης καταλήξαμε ότι τα διαστήματα εμπιστοσύνης για εκατοστημόρια για την υπό μελέτη μεταβλητή είναι έγκυρα. Το μόνο πρόβλημα που ανέκυψε στην εφαρμογή ήταν η ισχυρή ένδειξη της μη κανονικότητας των σφαλμάτων στα εκτιμηθέντα AR(1) και ARMA(1,1). Σε τι έκταση η μη κανονικότητα των σφαλμάτων επηρεάζει την εγκυρότητα των εκτιμηθέντων ασυμπτωτικών διαστημάτων εμπιστοσύνης σε δείγματα τουλάχιστον 50 παρατηρήσεων και με μέτρια επίπεδα αυτοσυσχέτισης χρήζει περαιτέρω διερεύνησης.

### **Ευχαριστίες**

Μία αρχική έκδοση της μελέτης αυτής παρουσιάστηκε στο 1<sup>ο</sup> Πανελλήνιο Συνέδριο Οικονομικής των Φυσικών Πόρων και του Περιβάλλοντος: Κλιματική Αλλαγή που πραγματοποιήθηκε στο Τμήμα Οικονομικών Επιστημών του Πανεπιστημίου Θεσσαλίας στο Βόλο 26-17 Μαρτίου 2014. Οι συγγραφείς ευχαριστούν τους συμμετέχοντες στο συνέδριο για τα εποικοδομητικά σχόλιά τους.

## ΠΑΡΑΡΤΗΜΑ

### Απόδειξη Λήμματος 1.

Η στοχαστική ανάλυση της σχέσης (1) εναλλακτικά γράφεται ως

$$X_t = \mu + \lim_{M \rightarrow \infty} \sum_{s=0}^{2M} \psi_s \varepsilon_{t-s} . \quad (A1)$$

Χρησιμοποιώντας τους μετασχηματισμούς  $\psi_s = \beta_{s-M}$  και  $\varepsilon_t = e_{t+M}$  (Fuller, 1996, σελ. 22), το δεξί σκέλος της (A1) παίρνει τη μορφή

$$X_t = \mu + \lim_{M \rightarrow \infty} \sum_{s=-M}^M \beta_s e_{t-s} = \mu + \sum_{s=-\infty}^{\infty} \beta_s e_{t-s} , \quad (A2)$$

με  $\{e_t\}$  ανεξάρτητες τυχαίες μεταβλητές κατανομήνες ως  $N(0, \sigma^2)$  και

$$\sum_{s=-\infty}^{\infty} |\beta_s| = \lim_{M \rightarrow \infty} \sum_{s=-M}^M |\beta_s| = \lim_{M \rightarrow \infty} \sum_{s=0}^{2M} |\psi_s| = \sum_{s=0}^{\infty} |\psi_s| < \infty . \quad (A3)$$

Η απόδειξη του τμήματος (α) του Λήμματος 1 ολοκληρώνεται χρησιμοποιώντας το θεώρημα 8.4.1 του Anderson (1971, σελ. 478), το οποίο επαναδιατυπώνεται ως εξής:

Εάν  $X_t = \mu + \sum_{s=-\infty}^{\infty} \beta_s e_{t-s}$  όπου  $\{e_t\}$  ανεξάρτητες μεταβλητές που έχουν την ίδια κατανομή με

με  $E(e_t) = 0$ ,  $E(e_t^2) = \sigma^2$ , και  $\sum_{s=-\infty}^{\infty} |\beta_s| < \infty$ , τότε το στατιστικό  $\sqrt{n}(\bar{X} - \mu)$  ασυμπτωτικά

κατανέμεται κανονικά με μέσο μηδέν και διακύμανση  $\sum_{s=-\infty}^{\infty} \gamma_s = \gamma_0 \sum_{s=-\infty}^{\infty} \rho_s$ .

Για την απόδειξη του τμήματος (β) του Λήμματος 1, θα χρησιμοποιήσουμε τις σχέσεις (A1)-(A3) και το θεώρημα 8.4.2 του Anderson (1971, σελ. 478):

Εάν  $X_t = \mu + \sum_{s=-\infty}^{\infty} \beta_s e_{t-s}$  όπου  $\{e_t\}$  ανεξάρτητες μεταβλητές που έχουν την ίδια κατανομή με

με  $E(e_t) = 0$ ,  $E(e_t^2) = \sigma^2$ ,  $E(e_t^4) = 3\sigma^4 + \kappa_4 < \infty$ , και  $\sum_{s=-\infty}^{\infty} |\beta_s| < \infty$ , τότε το διάνυσμα

$$\sqrt{n}[\hat{\gamma}_0 - \gamma_0 \quad \hat{\gamma}_1 - \gamma_1 \quad \hat{\gamma}_2 - \gamma_2 \quad \dots, \quad \hat{\gamma}_{n-1} - \gamma_{n-1}]'$$

ασυμπτωτικά ακολουθεί την  $n$ -διαστάσεων κανονική κατανομή με μέσο  $\mathbf{0}$  και μήτρα διακυμάνσεων-συνδιακυμάνσεων που προκύπτει από τη σχέση

$$\lim_{n \rightarrow \infty} n \text{Cov}(\hat{\gamma}_h, \hat{\gamma}_g) = \sum_{s=-\infty}^{\infty} [\gamma_{s+h} \gamma_{s+g} + \gamma_{s-g} \gamma_{s+h}] + \frac{\kappa_4}{\sigma^4} \gamma_h \gamma_g, \quad h = 0, 1, \dots, n-1, \quad g = 0, 1, \dots, n-1, \quad (A4)$$

όπου  $\hat{\gamma}_h = \frac{1}{n-h} \sum_{t=1}^{n-h} (X_t - \mu)(X_{t+h} - \mu)$ , και  $\kappa_4 = E(e_t^4) - 3\sigma^4$  είναι το 4<sup>ης</sup> τάξης cumulant.

Όταν  $e_t \sim N(0, \sigma^2)$  τότε  $\kappa_4 = 0$ , οπότε θέτοντας  $h = g$  στη σχέση (A4) το στατιστικό  $\sqrt{n}(\hat{\gamma}_0 - \gamma_0)$  ασυμπτωτικά θα έχει τη κανονική κατανομή με μέσο 0 και διακύμανση

$$2 \sum_{s=-\infty}^{\infty} \gamma_s^2 = 2\gamma_0^2 \sum_{s=-\infty}^{\infty} \rho_s^2 .$$



## ΑΝΑΦΟΡΕΣ

- Anderson, T.W. (1971). *The Statistical Analysis of Time Series*, A Wiley-Interscience Publication, John Wiley & Sons, Inc.
- Bland, J.M., Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8, 135-160
- Chakraborti, S., Li, J. (2007). Confidence interval estimation of a normal Percentile. *American Statistician* 61, 331-336
- Donner, A., Zou, G.Y. (2010). Closed-form confidence intervals for functions of the normal mean and standard deviation, *Statistical Methods in Medical Research* 21, 347-359
- Dyer, D.D., Keating, J.P., Hensley, O.H. (1977). Comparison of point estimators of normal percentiles. *Communications in Statistics, Simulation and Computations* B6, 269-283
- Efron, B., Tibshirani, R., J. (1993). *An Introduction to the Bootstrap*, New York, Chapman & Hall
- Fuller, W.A. (1996). *Introduction to Statistical Time Series* (2<sup>nd</sup> Ed.), A Wiley-Interscience Publication.
- Gibbons, J.D., Chakraborti, S. (2003). *Nonparametric Statistical Inference* (4<sup>th</sup> Ed.), New York, Marcel Dekker
- Halkos, G. E., Kevork, I. S. (2013a). Evaluating frequentist inferential approaches for optimal order quantities in the newsvendor model under exponential demand. *International Transactions in Operational Research* 20, 837-857.
- Halkos, G. E., Kevork, I. S. (2013b). Forecasting the optimal order quantity in the newsvendor model under a correlated demand. *MPRA Paper No. 44189*, Online at <http://ideas.repec.org/p/pra/MPRA/44189.html>
- Harvey A.C. (1993). *Time series models*. 2<sup>nd</sup> Edition, Prentice Hall, An imprint of Pearson Education Limited.
- Keating, J.P. (1983). Estimators of percentiles based on an absolute loss. *Communications in Statistics, Theory and Methods* A12, 441-447
- Keating, J.P., Mason, R.L., Balakrishnan, N. (2010). Percentiles estimators in location-scale parameter families under absolute loss. *Metrika* 72, 351-367
- Kevork, I. S. (2010). Estimating the optimal order quantity and the maximum expected profit for single period inventory decisions. *OMEGA, The International Journal of Management Science* 38, 218-227.
- Knight K. (2000). *Mathematical Statistics*. Taylor & Francis Ltd.